UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/647,709	08/25/2003	Senis Busayapongchai	60027.0303US01/BS030022	2462

23552 7590 05/09/2008

MERCHANT & GOULD PC
P.O. BOX 2903
MINNEAPOLIS, MN 55402-0903

EXAMINER	
SHAH, PARAS D	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
05/09/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/647,709	BUSAYAPONGCHAI, SENIS	
	Examiner	Art Unit	
	PARAS SHAH	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *25 March 2008*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-11,15-17 and 22-24* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-11,15-17 and 22-24* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. This Office Action is in response to the RCE filed on 03/25/2008. Claims 1-11, 13, 15-17, and 22-24 remain pending with. All mentioned claims have been examined. The Applicants' amendment and remarks have been carefully considered but they do not place the case in condition for allowance.

2. All previous objections and rejections directed to the Applicant's disclosure and claims not discussed in this Office Action have been withdrawn by the Examiner.

Response to Amendments and Arguments

3. Applicant's arguments (pages 7-10) filed on 03/25/2008 with regard to claims 1-11, 13, 15-17, and 22-24 have been fully considered but they are moot in view of new grounds for rejection. Thus, the prior art reference of Valles (US 2004/0083092) has been removed and the prior art reference of Richardson (US 5,999,896) has been applied.

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically teach or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

5. Claims 1, 3-6, 10, 11, 15, 16, and 22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Crepy *et al*. (US 6,622,121) in view of Richardson *et al.* (US 5,999,896) in view of Raud *et al.* (US 6,125,341).

As to claim 1, Crepy *et al.* teaches a method for testing and improving the performance of a speech recognition engine, comprising:

loading into a memory location one or more words, phrases or utterances of plural grammar types (see col. 2, lines 64-66 and col. 3, lines 36-39) (e.g. The words inputted from the text contains various types of words and thus are of plural grammar types (i.e. subject or domain).);

identifying one or more of the words, phrases or utterances for recognition by a speech recognition engine (see col. 3,lines 36-39 and col.3,lines 40-43) (e.g. It is seen that the reference text, which consists of words are identified and will be passed to the speech recognition);

extracting the one or more words, phrases or utterances in a selected grammar sub-tree via a vocabulary extractor module and, passing the extracted one or more identified words, phrases or utterances to a text-to-speech conversion module that provides an audio formatted pronunciation of each word, phrase, or utterance (see col.3, lines 36-46 and col. 1, line 65-67) (e.g. The extracted words come from the reference text, which is then fed into the text to speech engine. An audio representation is produced as a result of the conversion of text into speech.);

passing the audio pronunciation of each of the identified one or more words, phrases or utterances, from the text-to- speech conversion module to the speech recognition engine (see col. 4, lines 59-65 and Figure 4, elements, 404 and 406).;

creating a recognized word, phrase or utterance for each audio pronunciation passed to the speech recognition engine (see col. 4, lines 59-65) (e.g. It is seen that the words are recognized from the audio file and then compared.); and

analyzing each recognized word, phrase or utterance created by the speech recognition engine to determine how closely each created recognized word, phrase or utterance approximates the respective audio pronunciation from which each created recognized word, phrase or utterance is derived (see col. 4, lines 65-col. 5, lines 11) (e.g. It is seen that a comparison is done with regards to the recognized words and the actual words using the WER calculation.)

However, Crepy *et al.* does not specifically teach the categorizing by the identified spoken words by grammar type where same utterances are grouped together in a grammar sub-tree and selection of a particular grammar sub-tree.

Richardson *et al.* does teach use of spoken words (see col. 3, lines 39-42, voice recognizer allows user to input voice for conversion into text)

categorizing the identified one or more words, phrases or utterances (see col. 3, lines 45-57, confusable words are identified and categorized based on a confusable word table) by grammar type (see Figure 4, and col. 4, lines 37-39,

the confusable words are separated by type of confusable word pair, alphabetically) whereby all words, phrases or utterances of a same grammar type are grouped together in a grammar sub-tree (see Figure 4, for example, the word their, the words “there” and “they're” are grouped together as other possible words for grammar type “their")

selecting a particular grammar sub-tree (see col. 5, lines 47-59, user is presented with choices of a grammar sub-tree for grammar of confusable word that was identified (see Figure 7))

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the improving of speech recognition as taught by Crepy *et al.* with the inclusion of categorizing words according to a specific grammar as taught by Richardson. The motivation to have combined the references involves the ability to resolve commonly confused words (See Richardson *et al.* col. 1, lines 51-53).

However, Crepy *et al.* in view of Richardson *et al.* do not specifically teach the assignment of confidence score for each utterance, phrase, or word

Raud *et al.* teaches assigning a confidence score to each utterance, phrase or word (see col. 6, lines 8-21).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the improving of speech recognition as taught by Crepy *et al.* in view of Richardson *et al.* with the inclusion of assigning confidence score as taught by Raud *et al.* The motivation to have combined the

references involves the ability determine if the current vocabulary is appropriate for recognizing words and to determine of a word is properly recognized (see Raud *et al.* col. 6, lines 8-13).

As to claim 15, Crepy *et al.* in view of Richardson in view of Raud teach all the claimed limitations as applied to claim 1 above

Furthermore, Richardson teaches a plurality of grammar sub-trees are grouped together to form a grammar tree containing all of the one or more words, phrases, or utterances (see Figure 4) (e.g. The figure shows that a plurality of confusable words of different grammar types is shown with possible intended words or sub-trees that are linked to the candidate confusable word.)

As to claim 16, Crepy *et al.* in view of Richardson in view of Raud teach all the claimed limitations as applied to claim 1 above.

Furthermore, Crepy teaches the use of a speech recognition engine (see Crepy *et al.*, Figure 4, element 406)

Furthermore, Richardson teaches the identifying of an utterance includes selecting the grammar sub-tree containing the one or more words, phrases, or utterances (see col. col. 4, lines 57-61, parser identifies confusable words by relating to a table).

As to claim 3, Crepy *et al.* in view of Richardson *et al.* in view of Raud *et al.* teaches all the claimed limitations as applied to claims 1 and 2 above.

Furthermore, Raud *et al.* teaches the assigning of confidence score to each recognized utterance based on a confidence level associated with the utterance based on prior speech recognition engine training (see Raud *et al.* col. 6, line 8)(e.g. It is obvious that the confidence score is compared based on a threshold for recognition accuracy (see col. 6, lines 23-31).

As to claims 4 and 10, Crepy *et al.* in view of Richardson *et al.* in view of Raud *et al.* teaches all the claimed limitations as applied to claims 1 and 3 above.

Furthermore, Raud *et al.* teaches the determination being made of whether the recognized utterance is the same as the utterance derived by the speech recognition engine based on prior speech recognition training confidence level (see Raud *et al.*, col. 4, lines 33-35)) (e.g. It should be noted that there is a vocabulary used for checking if there is a match. An initial vocabulary is used, then other vocabularies are used for subsequent words not found or recognized using the initial vocabulary (see col. 5,lines 46-56). It is inherent that the words from the vocabulary and the words from the utterance are matched for similarity).

As to claims 5 and 11, Crepy *et al.* in view of Richardson *et al.* in view of Raud *et al.* teach all the claimed limitations as applied to claims 1 and 2 above.

Furthermore, Raud *et al.* teaches if the confidence score exceeds an acceptable level designating the recognized utterance as accurately recognized by the speech recognition engine (see Raud *et al.* col. 5, lines 18-30).

As to claim 6, Crepy *et al.* in view of Richardson *et al.* in view of Raud *et al.* teaches all the claimed limitations as applied to claims 1, 2, and 5 above.

Furthermore, Raud *et al.* teaches if the confidence score less than a certain value, a modification is made to the speech recognition engine to recognize the word (see col. 6, lines 8-31) (e.g. If the confidence level is less than a value, the system requests verification from a user or asks a question to remove any ambiguity. This is seen as a modification to the speech recognition engine to interpret the utterance. Further, other vocabularies are used to determine whether an increase in performance can be obtained.).

As to claim 22, Crepy *et al.* teaches a method for testing and improving the performance of a speech recognition engine, comprising:

identifying one or more of the words, phrases or utterances for recognition by a speech recognition engine (see col. 3,lines 36-39 and col.3,lines 40-43) (e.g. It is seen that the reference text, which consists of words are identified and will be passed to the speech recognition);

creating and passing the audio pronunciation of each of the identified one or more words, phrases or utterances, from the text-to- speech conversion

module to the speech recognition engine that provides an audio formatted pronunciation of each of the identified words, phrases, or utterances to the speech recognition engine (see col. 4, lines 59-65 and Figure 4, elements, 404 and 406) (e.g. It is seen from the cited section that an audio version is created of the input speech and passed to the speech recognition engine.);

deriving a recognized word, phrase or utterance for each audio pronunciation passed to the speech recognition engine; (see col. 4, lines 65-col. 5, lines 11) (e.g. It is seen that a comparison is done with regards to the recognized words and the actual words using the WER calculation.)

However, Crepy *et al.* does not specifically teach the categorizing by a grammar type where same utterances are grouped together in a grammar sub-tree.

Richardson *et al.* does teach use of spoken words (see col. 3, lines 39-42, voice recognizer allows user to input voice for conversion into text)

categorizing the identified one or more words, phrases or utterances (see col. 3, lines 45-57, confusable words are identified and categorized based on a confusable word table) by grammar type (see Figure 4, and col. 4, lines 37-39, the confusable words are separated by type of confusable word pair, alphabetically) whereby all words, phrases or utterances of a same grammar type are grouped together in a grammar sub-tree (see Figure 4, for example, the word their, the words “there” and “they're” are grouped together as other possible words for grammar type “their")

selecting a particular grammar sub-tree (see col. 5, lines 47-59, user is presented with choices of a grammar sub-tree for grammar of confusable word that was identified (see Figure 7))

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the improving of speech recognition as taught by Crepy *et al.* with the inclusion of categorizing words according to a specific grammar as taught by Richardson. The motivation to have combined the references involves the ability to resolve commonly confused words (See Richardson *et al.* col. 1, lines 51-53).

However, Crepy *et al.* in view of Richardson *et al.* do not specifically teach the assignment of confidence score for each utterance, phrase, or word.

Raud *et al.* teaches the assigning a confidence score to each utterance, phrase or word (see col. 6, lines 8-21) based on prior training of the speech recognition engine to recognize similar or same words, phrases or utterances as t-he-each derived recognized word, phrase or utterance (see Raud *et al.*, col. 4, lines 33-35) (e.g. It should be noted that there is a vocabulary used for checking if there is a match. An initial vocabulary is used, then other vocabularies are used for subsequent words not found or recognized using the initial vocabulary (see Raud *et al.*, col. 5,lines 46-56). It is inherent that the words from the vocabulary and the words from the utterance are matched for similarity). and.

if the confidence score is less than an acceptable threshold, modifying the speech recognition engine to recognize with higher accuracy the word, phrase or

utterance from which the derived recognized word, phrase or utterance is derived higher accuracy (see col. 5, lines 31-38 and col. 6, lines 22-51).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the improving of speech recognition as taught by Crepy *et al.* and Richardson *et al.* with the inclusion of assigning confidence score as taught by Raud *et al.*. The motivation to have combined the references involves the ability determine if the current vocabulary is appropriate for recognizing words and to determine of a word is properly recognized (see col. 6, lines 8-13).

6. Claim 7 is rejected under 35 U.S.C. 103(a) as being unpatentable over Crepy *et al.* in view of Richardson *et al.* and Raud *et al.* as applied to claim 5 above, and further in view of Bickley *et al.* (US 7,013,276).

As to claims 7, Crepy *et al.*, Richardson *et al.* and Raud *et al.* teach improving the performance of a speech recognition engine.

However, Crepy *et al.*, Richardson *et al.* and Raud *et al.* do not specifically teach the notification to a developer when the score is lower than a threshold value.

Bickley *et al.* teaches a alert mechanism for words that are similar and are subject to confusion (see col. 10, lines 63-65) from threshold calculation (see col. 10, lines 38-40).

It would have been obvious to one of ordinary skilled in the art to modify the speech recognition performance methods as taught by Crepy *et al.*, Richardson *et al.* and Raud *et al.* with the use of a notification sent to a software developer when value is below threshold as taught by Bickley *et al*. The motivation to combine these references involves the distinguishing between similar words, which may not be recognized by speech recognition engines (see Bickley *et al.* col. 2, line 27-36).

7. Claim 17 is rejected under 35 U.S.C. 103(a) as being unpatentable over Crepy *et al.* in view of Richardson *et al.* in view of Raud as applied to claim 1 above, and further in view of Kennewick *et al.* (2004/0044516).

As to claim 17, Crepy *et al.* in view of Richardson *et al.* in view of Raud teach all the claimed limitations as applied to claims 1.

Furthermore, Crepy *et al.* teaches the creating of a recognized word, phrase, or utterance for each respective audio pronunciation includes converting each respective audio pronunciation from an audio format to a digital format by the speech recognition engine (see Crepy *et al.*, col. 4, lines 56-64). (e.g. It is seen that the audio form of the file is converted into the digital form. The words contain an implied pronunciation of the words.).

However, Crepy *et al.* in view of Richardson *et al.* in view of Raud do not specifically teach the analyzing phonetically each respective audio pronunciation of each of the one or more recognized word, phrase or utterance.

Kennewick *et al.* does teach

the analyzing phonetically each respective audio pronunciation of each of the one or more recognized word, phrase or utterance (see [0151]).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the improving of speech recognition as taught by Crepy *et al.* and Richardson *et al..* with the inclusion of analyzing the phonetics of each audio pronunciation. The motivation to have combined the references involves the add pronunciations not present in the dictionary in order to increase speech recognition accuracy and learning (see Kennewick *et al.*, [0151]).

8. Claims 8 and 23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Crepy *et al.*, Richardson *et al.* and Raud *et al.* as applied to claims 6 and 22 above, and further in view of Kennewick *et al.* (US 2004/0044516).

As to claims 8 and 23, Crepy *et al.*, Richardson *et al.* and Raud *et al.* teach all the claimed limitations as applied to claims 1, 5, and 6 above and claim 22. Furthermore, Raud *et al.* teaches the assigning of a confidence score and if less than a threshold, obtaining an acceptable confidence score upon next pass through the engine (see col. 7, lines 20-25)

However, Crepy *et al.*, Richardson *et al.* and Raud *et al.* do not specifically teach the altering of the audio pronunciation with the confidence score less than an acceptable threshold.

Kennewick *et al.* does teach the altering of audio pronunciation of the word, phrase, or utterance associated with the confidence score that is less than an acceptable confidence score threshold level such that the altered audio pronunciation obtains an acceptable confidence score upon next pass through the speech recognition engine (see [0151]). (e.g. The speech recognition engine is adaptive based on the confidence levels and the pronunciation of the word recognized.)

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the improving of speech recognition as taught by Crepy *et al.*, Richardson *et al.* and Raud *et al.* with the inclusion of altering the audio pronunciation of the recognized word as taught by Kennewick *et al*. The motivation to have combined the references involves the ability to improve the accuracy of the speech recognition engine as well as the ability for the speech recognition engine to learn with time (see Kennewick *et al*., [0151]).

9. Claims 9 and 24 are rejected under 35 U.S.C. 103(a) as being unpatentable over Crepy *et al.* in view of Richardson *et al.* and in view of Raud *et al.* as applied to claims 6 and 22 above, and further in view of Roberts *et al.* (US 6,999,930).

As to claims 9 and 24, Crepy *et al.* in view of Richardson *et al.* in view of Raud *et al.* teach all the claimed limitations as applied to claims 1, 5, and 6 above and claim 22. Furthermore, Raud *et al.* teaches the use of a confidence score (see col. 6, lines 23-31).

Crepy *et al.* in view of Richardson *et al.* in view of Raud *et al* do not specifically teach the reduction of the confidence threshold level.

However, Roberts does teach the reduction of the confidence score threshold level (see col. 10, lines 50-60).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the improving of speech recognition as taught by Crepy *et al.* in view of Richardson *et al.* in view of Raud *et al.* with the inclusion of altering the reducing the acceptable confidence sore threshold level as taught by Roberts *et al*. The motivation to have combined the references involves the ability to generate more potential matches even when the confidence level is low (see Roberts *et al.,* col. 10, lines 57-60).

Conclusion

10. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Dragosh *et al.* (US 6,856,960) is cited to disclose the selection of grammars, which consists of sub-grammars for use in TTS and speech recognition. Rusnak *et al.* is cited to disclose a domain specific concatenate audio based on domains.

11. Any inquiry concerning this communication or earlier communications from the examiner should be directed to PARAS SHAH whose telephone number is (571)270-1650. The examiner can normally be reached on MON.-THURS. 7:00a.m.-4:00p.m. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Patrick Edouard can be reached on (571)272-7603. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Paras Shah/
Examiner, Art Unit 2626

04/28/2008

/Patrick N. Edouard/

Supervisory Patent Examiner, Art Unit 2626